EVALUATION OF METHODS FOR INTEGRATING MONITORING AND MODELLING DATA FOR REGULATORY AIR QUALITY ASSESSMENTS

Angela Ball, Richard Hill and Peter Jenkinson

Westlakes Scientific Consulting Ltd, The Princess Royal Building, Westlakes Science and Technology Park, Moor Row, Cumbria, CA24 3LN, UK.

Abstract: Measured and modelled SO_2 concentration data from Kincaid, Illinois (USA) and the Aire Valley (UK) were used to evaluate four data assimilation methods to determine their effectiveness in calibrating modelled air concentrations. The data assimilation methods included two simple methods (linear regression and simple ratio), which resulted in a global adjustment of the modelled concentration field and two complex methods (kriging of the ratio and kriging of the residual), which resulted in a spatially varying calibration of the modelled concentration field. Prior to the analysis, the measured data were 'sector-corrected' to remove the influence of external sources to ensure a direct comparison between the calibration methods. A cross-validation technique, applying standard model evaluation statistics, was used to assess the performance of each calibration method. The simple ratio method provided the most accurate model calibration for both the Kincaid and the Aire Valley data sets by minimising the normalised mean square error and mean bias and maximising the fraction of modelled predictions within a factor of two of measured predictions. The linear regression method performed to a similar level when using a high number of data points, although the performance declined dramatically when just two monitoring points were used for calibration. The more complex kriging methods were found to be less effective than the simple methods, despite offering a spatially varying model calibration. The analysis of the Kincaid data set suggests that between 10 and 15 monitoring points may be necessary for the optimum calibration of a modelled concentration field using the simple calibration methods reviewed in this study.

Key words: *data assimilation, model calibration, sector-correction, kriging, cross-validation.*

1. INTRODUCTION

Ambient monitoring and air quality modelling are the tools most frequently used to assess the impact of industrial emissions to air. The use of either modelling or monitoring alone may introduce a high degree of uncertainty into air quality impact assessments. However, it is anticipated that by applying monitoring and modelling integration techniques, the strengths of both technologies can be complemented and the uncertainties associated with each technique reduced.

Data assimilation is a procedure that combines measured data with modelled data to provide a better estimation of the quantity of interest. In the context of air quality assessment, data assimilation has several purposes:

- To calibrate retrospective modelling assessments and determine realistic impacts in areas where monitoring was not conducted.
- To calibrate prospective modelling assessments using measurement data to normalise modelled data.
- To provide improved estimates of source attribution and estimates of source strength.
- To optimise the design of monitoring networks.

Data assimilation methods may be described as simple or complex. Simple methods involve a global adjustment of the modelled concentration field, with no spatial or temporal dimension. Complex methods may incorporate either a spatially or temporally varying adjustment of the modelled concentration field, or for the most complex methods, both adjustments combined. However, the use of these methods for the calibration of air quality assessments has been limited due to the complexity of the data assimilation methods involved. In addition, the effectiveness of these methods depends on an adequate spatial and temporal coverage of monitoring data.

Examples of data assimilation methods currently in use in air quality management include the linear regression method adopted in Local Air Quality Management Technical Guidance TG(03) (Defra, 2003) and the standard calibration procedure routinely applied to output data from the FRAME (Fine Resolution Atmospheric Multi-pollutant Exchange) model (Dore *et al.*, 2005). This paper examines four data assimilation methods, both simple and complex, to determine their effectiveness in calibrating modelled predictions of air quality impacts from industrial point sources.

2. DATA ASSIMILATON METHODS

The calibration methods compared in this study include: 1) linear regression; 2) simple ratio; 3) kriging of the ratio; and 4) kriging of the residual. Both the linear regression and simple ratio methods perform a global adjustment of the modelled concentration field. By contrast, kriging is a more powerful data assimilation tool, which may be used to generate spatial information by interpolating between points of known value to produce a spatially varying adjustment of the modelled concentration field. For Method 1, a linear regression model was derived from a comparison between measured and modelled data. The slope and intercept were used to provide appropriate correction factors by which modelled concentrations were adjusted. For Method 2, the mean ratio of measured-to-modelled concentrations was used to perform a global adjustment to the modelled concentration field. For Method 3, a kriged surface was generated from the ratios of modelled-to-measured concentrations using ordinary kriging and subsequently used to calibrate the modelled concentration field. For Method 4, a kriged surface was generated from the residuals (the difference between modelled and measured concentrations) using ordinary kriging and subsequently used to calibrate the modelled concentration field.

3. KINCAID SO_2 DATA SET

The Kincaid SO_2 model evaluation data set contains 6024 hours of measurement data, collected during 1980-1981, from 28 monitoring sites around the Kincaid power plant in Illinois, USA. An interpolated plot of measured average SO_2 concentration in the Kincaid network, overlain with directional "pollution roses", is shown in Figure 1. The Kincaid data set was used for the validation of the AERMOD dispersion modelling system, therefore modelled concentration data were obtained from the US Environmental Protection Agency website (http://www.epa.gov/scram001/dispersion_prefrec.htm).

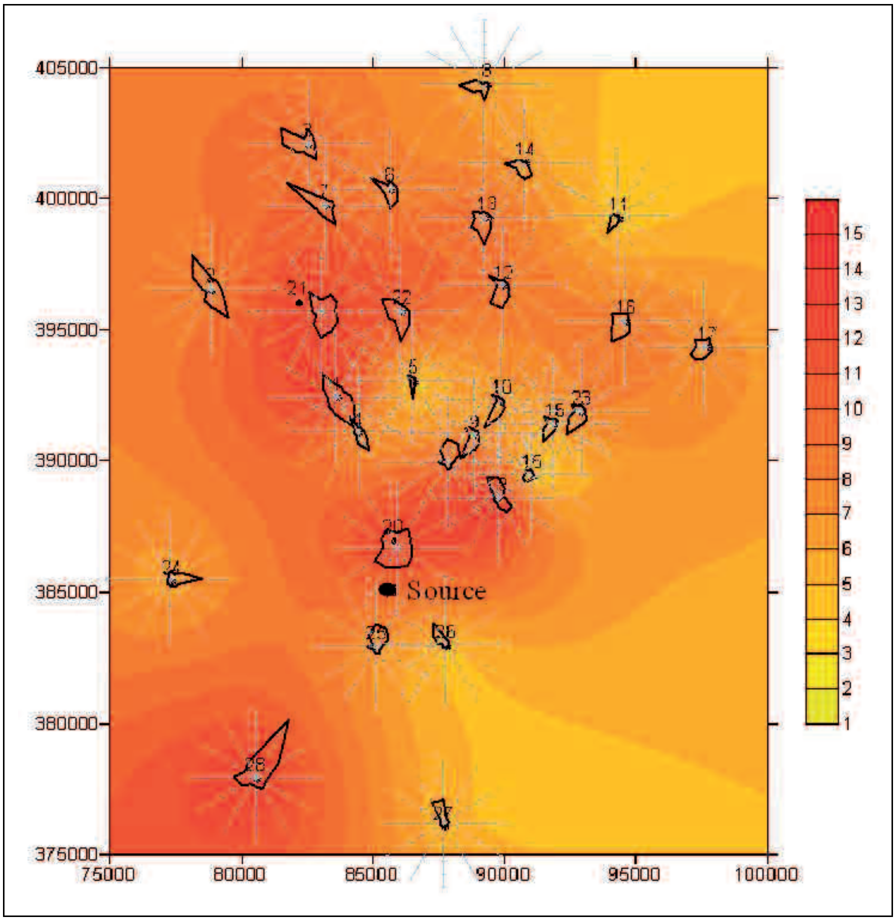

Figure 1. Interpolated plots of average SO_2 concentration (μgm^{-3}) in the Kincaid network, with superimposed annual average pollution rose diagrams, for data from 1980/81. Pollution rose segments are split into 30˚ sectors with intervals of 5 μgm^{-3}.

4. AIRE VALLEY DATA SET

The Aire Valley 2003 data set contains 15-minute mean SO_2 data from six monitoring sites, measuring the combined impact of emissions from three coal-fired power stations in West Yorkshire, UK. An interpolated plot of measured average SO_2 concentration in the Kincaid network is shown in Figure 2. Modelled concentration data were generated using ADMS 3.3 with no background correction. Hourly sequential meteorological data were supplied by the Met Office for Linton-on-Ouse (Grid reference: 4492 4613), approximately 40 km north of the Aire Valley. A roughness length of 20 cm was applied to the model domain, consistent with the roughness length used in the Joint Environment Programme (JEP) modelling scenarios for the Aire Valley (Brooke *et al.*, 2003). The land use at Linton-on-Ouse was considered similar to that in the Aire Valley, therefore the same roughness length was assigned to the meteorological site as for the model domain. A higher than expected frequency of very calm conditions (18%) was recorded in 2003 and consequently concentrations were not modelled for these hours.

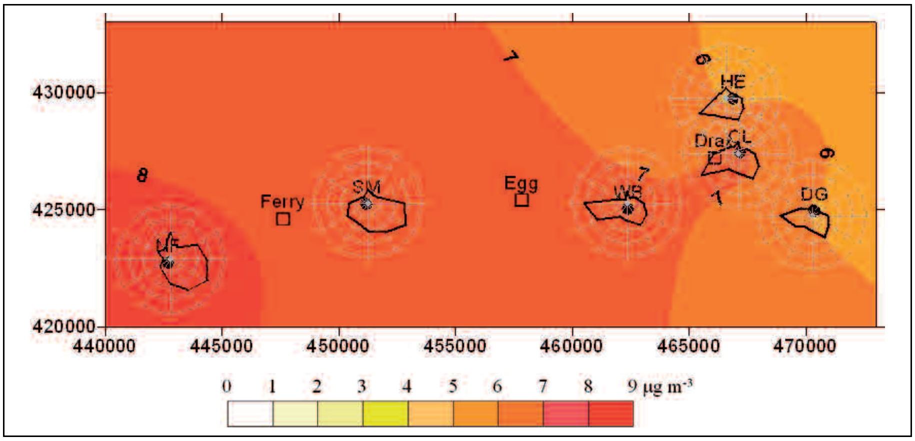

Figure 2. Interpolated plots of 2003 annual average SO_2 concentration (μgm^{-3}) in the Aire Valley, with superimposed annual average pollution rose diagrams for six monitoring site locations. Abbreviations; 'Ferry', 'Egg' and 'Drax' denote power stations; Ferrybridge (1 and 2), Eggborough and Drax, respectively. Abbreviations; NF, SM, WB, HE, CL and DG denote monitoring sites; North Featherstone, Smeathalls Farm, West Bank, Hemingbrough Landing, Carr Lane and Downes Ground, respectively. Pollution rose segments are split into 30° sectors with intervals of 5 μgm^{-3}.

5. SECTOR-CORRECTION OF MONITORING DATA

For both the Aire Valley and Kincaid data sets, the measured concentrations were influenced by sources external to the model domain. For example, Figure 2 shows annual average SO_2 pollution rose diagrams for monitoring sites in the Aire Valley, superimposed onto an annual average SO_2 contour plot. It is evident from the shape of the pollution rose, that a source to the south east of the Aire Valley is influencing measured concentrations. In order to analyse the performance of the calibration methods it was necessary to remove the influence of external sources by 'sector-correcting' the measurement data. This was achieved by calculating the mean concentration (and 99.9th percentile) using measurements recorded when the plume was orientated on a source-receptor trajectory within given wind arc sector limits. To determine the optimum sector size, mean concentrations (and 99.9th percentile) were calculated using wind arcs ranging from 5° to 100°, increasing in increments of 5°. The Root Mean Square Error (RMSE) and the regression r^2 of the modelled vs. sector-corrected measured data were calculated for each sector. The optimum sector size is determined where the RMSE was minimised, the r^2 maximised and where RMSE/R^2 is closest to zero, as shown in Figure 3. The optimum sector size for the Kincaid and Aire Valley data sets was found to be 25° and 60°, respectively.

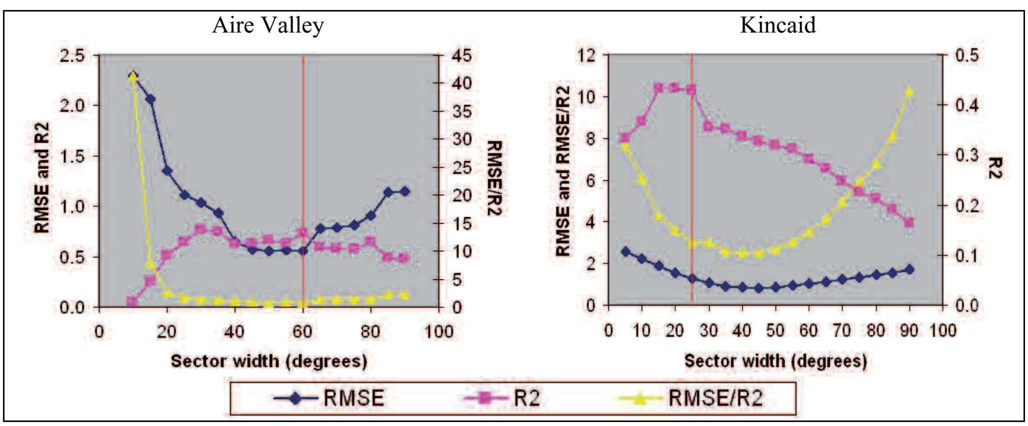

Figure 3. Sector correction of the Aire Valley and the Kincaid data sets, the vertical red line shows the optimal sector size.

6. ANALYSIS OF DATA ASSIMILATION METHODS

The calibration methods were analysed using a cross-validation technique known as 'jack-knifing', which involves the systematic re-calibration of modelled data, leaving out one monitoring site at a time from the measurement data set. The retained monitoring sites are used to conduct the calibration and the removed monitoring sites used to test the method by comparison with the calibrated model results. However, rather than removing a single monitoring site, an increasing number of sites were removed to determine the density of monitoring sites necessary for optimum model calibration. This resulted in numerous 'calibration scenarios' for each data assimilation method. For each calibration scenario (using data from *n* number of monitoring sites), one hundred random combinations of *n* monitoring sites (including repetitions where necessary) were generated to produce one hundred separate calibrations. The random selection of monitoring site combinations means that any repetition did not skew the subsequent analysis. The data assimilation methods were analysed using standard model evaluation statistics; normalised mean square error (NMSE), mean bias (MB) and the percentage of modelled predictions within a factor of 2 (F2) of measured predictions, expressed as a value between 0 and 1.

7. RESULTS

Figures 4 (a - f) show summary statistics (NMSE, F2 and MB) comparing measured vs. calibrated modelled annual mean SO_2 concentrations for Kincaid and the Aire Valley. Shown below each diagram, in bold, is the summary statistic generated by comparing measured vs. uncalibrated model values.

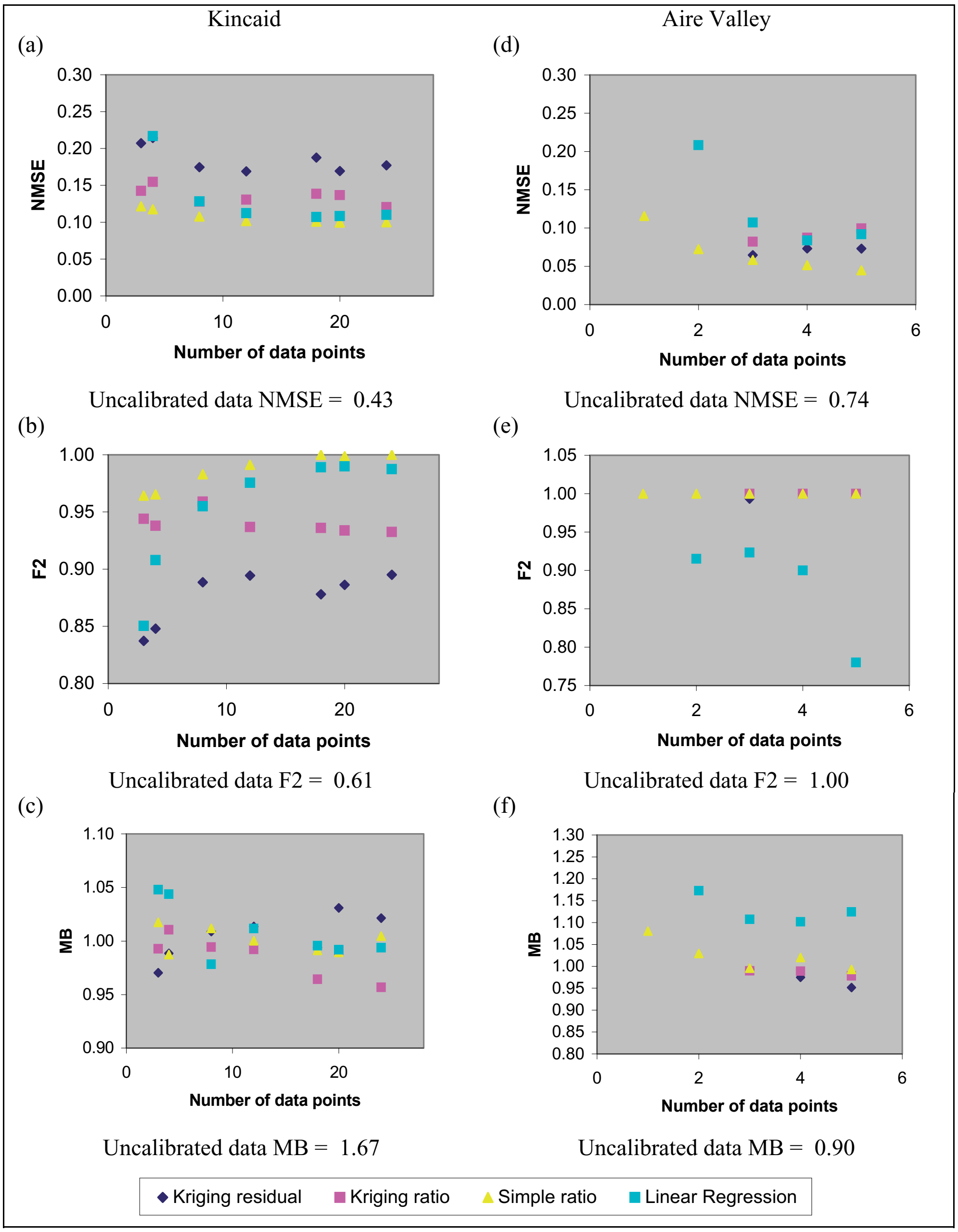

Figure 4. Statistical analysis of the performance of each calibration method when applied to the annual average SO_2 concentration data for Kincaid and the Aire Valley. The x-axis label denotes the number of monitoring points used to conduct the calibration. The summary statistics generated by comparing measured vs. uncalibrated model data are shown below each diagram for comparison.

For the Kincaid data set, the summary statistics show that all four calibration methods have improved the consistency between modelled and measured concentrations when compared with the uncalibrated data. Figures 4 (a) and (b) indicate that as the number of data points used to perform the calibration increases, the F2 statistic increases, with the exception of the kriging ratio method in Figure 4 (b) which maintains a relatively constant F2 value of 0.93-0.96. However, the integration of data using more than 10 to 15 monitoring points was found to provide no further reduction in cross-validation error terms. Overall, the simple ratio method appears to perform best as this method constantly produces the lowest NMSE and the highest F2 statistics, regardless of the number of data points used to conduct the calibration.

The summary statistics for the Aire Valley data set also show that all four calibration methods have improved the consistency between modelled and measured concentrations when compared with the uncalibrated data. However, Figures 4 (d, e and f) show that the linear regression method performs poorly when only 2 data points are used to perform the calibration. The NMSE statistic in Figure 4 (d) indicates that the simple ratio is the best performing calibration method for the Aire Valley data. However, it is difficult to tell how the kriging methods perform because a minimum of three data points is needed to produce a kriged surface and the maximum number of data points available from the Aire Valley data set is five. This is a consistent problem with the Aire Valley data set and limits any comparisons with Kincaid. For most calibration methods, NMSE and MB statistics show that as the number of data points used to perform the calibration increases, the calibration performance is improved. However, Figure 4 (e) illustrates that this is not the case for the linear regression method when used to calculate the annual average concentration. This may be because an outlier is affecting the linear regression properties and therefore producing a poor calibration. The analysis of the Aire Valley data indicates that the calibration methods are optimised using five monitoring points (the maximum number available). Although Figures 4 (d) and (e) seem to show stabilising summary statistics for the simple ratio method, it is difficult to ascertain whether increasing the number of monitoring sites, beyond the six already established, would further improve the performance of the calibration methods.

8. CONCLUSIONS

Four calibration methods have been analysed to determine their effectiveness in calibrating modelled predictions of air quality impacts from industrial point sources. The simple ratio method proved best using both the Kincaid data set and the Aire Valley data set. The linear regression method performed on a similar level to the simple ratio method when using a high number of data points but the performance of this method declined dramatically with the use of only two monitoring points. This is due to the poor definition of the linear regression when using a low number of data points. Overall, the more complex kriging methods appear to be less effective than the simple methods. The analysis of the Kincaid data set suggests that the number of monitoring points required to enable optimum calibration is between 10 and 15.

Statistical optimisation of model calibration is only one consideration in determining the number of monitoring sites for regulatory purposes. Nonetheless, if the costs of monitoring decline through developments in low cost, high-resolution sensor technologies, these larger data fields may become achievable. It should be noted that data assimilation is designed to reduce uncertainty in model outputs caused by systematic errors due to model formulation and is of limited use when uncertainty is due to random errors e.g. the stochastic nature of atmospheric turbulence.

Acknowledgements: *This work was funded by the Environment Agency under the umbrella project 'Methods to measure and model ambient air pollution at local scales (<10km)'. A full version of the report titled 'Integration of air quality modelling and monitoring methods: review and applications' is available from the Environment Agency on-line catalogue at http://publications.environment-agency.gov.uk/epages/eapublications.storefront.*

REFERENCES

Brooke, D., V. Lucas and S. Bethan, 2003: A comparison of AERMOD with ADMS 3.1 and measured data from Yorkshire and the Lower Trent Valley in 1998 and 1999. RWE Innogy JEP Report, ENV/EEA/110/03.

Defra, 2003: Local Air Quality Management Technical Guidance LAQM TG(03). Department for Environment, Food and Rural Affairs Publication, PB7514.

Dore, A.J., M. Vieno, W. J. Bealey and M.A. Sutton, 2005: Source Attribution, Report as part II of the project: Source Attribution and critical loads assessment for Special Areas of Conservation and Special Protection Areas in the UK. Centre for Ecology and Hydrology (Edinburgh Research Station), Bush Estate, Penicuik, Midlothian, EH26 0QB, UK.